Markov Decision Theoretic Pilot Allotment & Receive Antenna Selection

Reuben George Stephen

SPC Lab,
Indian Institute of Science,
Bangalore-560012
(reubengs@ece.iisc.ernet.in)

March 10, 2012

Antenna Selection (AS)

- Popular technique to reduce hardware costs
- Uses fewer RF chains than actual number of antenna elements
- Process signals from a dynamically selected subset of antennas only
- Achieves same diversity order as a full-complexity system [Molisch and Win, 2004]

Existing Work

- Several algorithms proposed assuming perfect CSI at the receiver ([Wang et al., 2010] & references therein)
- In practice, CSI needs to be acquired
- Imperfect CSI $\Rightarrow$ inaccurate selection, imperfect data decoding $\Rightarrow$ increased SEP [Kristem et al., 2010]
- But, AS achieves same full diversity order as with perfect CSI even with channel estimation errors [Gucluoglu and Panayirci, 2008]
- Concentrate on single receive antenna selection

Motivation

- Consider packet reception, time divided into frames
- Correlated time-varying channel $\Rightarrow$ could exploit correlation to aid in antenna selection decision
- With pilot-based training, prior information can also aid in deciding how accurately a channel at a particular antenna should be estimated
- Link-level error checks on data packets $\Rightarrow$ provides additional info on channel state at selected antenna $\Rightarrow$ can again be used in future pilot allotment/antenna selection decisions.

System Model I

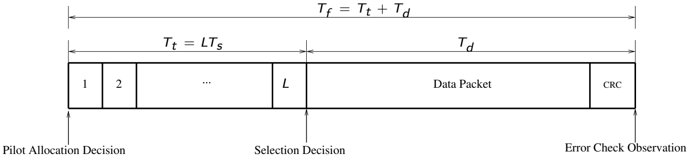

Figure: Frame structure for training & data reception

- 1 transmit antenna, N receive antennas, 1 RF chain
- Channel at antenna i, $h_i[k]$, constant for whole frame k, but correlated across frames
- Receiver can decide how many pilots to receive with antenna i in frame k, $\ell_i[k]$

System Model II

- Allocation of $\ell_i[k]$ would influence selection decision and hence, the throughput

Objective

In each k choose $\ell_i[k]$ $\forall i$, select $n \in \{1, \ldots, N\}$, to maximize expected long-run throughput

- Problem can be modeled as a partially observable Markov decision process (POMDP)

Markov Decision Process I

- Model for agent interacting with world
- No uncertainty about current state

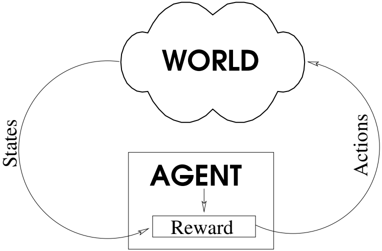

Figure: MDP

Markov Decision Process II

$\langle \mathcal{S}, \mathcal{A}, \mathcal{T}, R \rangle$

$\mathcal{S}$ states
$\mathcal{A}$ actions
$\mathcal{T} : \mathcal{S} \times \mathcal{A} \to \Pi(\mathcal{S})$ state transition function
$R : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ reward function

- Given $s \in \mathcal{S}$ and $a \in \mathcal{A}$ at t, s_{t+1} and R_{t+1} independent of all past states and actions
- Objective: Maximize reward over finite/infinite horizon
- Policy $\pi_t : \mathcal{S} \to \mathcal{A}$

POMDP I

- Agent cannot determine current state with complete reliability

$\langle \mathcal{S}, \mathcal{A}, \mathcal{T}, R, \Omega, \mathcal{O} \rangle$

MDP $\langle \mathcal{S}, \mathcal{A}, \mathcal{T}, R \rangle$
Ω observations
$\mathcal{O} : \mathcal{S} \times \mathcal{A} \rightarrow \Pi(\Omega)$ observation function

POMDP II

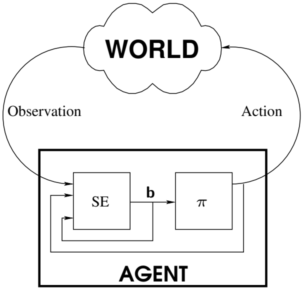

Figure: POMDP agent

- Belief state $\mathbf{b} \in \Pi(\mathcal{S})$, sufficient statistic for past history and initial belief state
- Policy π is now a function of $\mathbf{b}$
- Optimal policy is solution of continuous space "belief MDP"

Simplified Channel Model I

- For simplicity, assume 2-state channel with $h_i[k] \in \{h_0, h_1\}$, $|h_0| \ll |h_1|$, and $h_0, h_1 \in \mathbb{C}$ known to receiver
- Assume that successful packet reception depends only on true channel state, rather than receiver's estimate.
- $\mathbf{p}_i = \sqrt{\frac{E_p}{L}}[1, \ldots, 1]^H \in \mathbb{C}^{\ell_i}$, vector of pilot symbols
- $\mathbf{y}_i = [y_1, \ldots, y_{\ell_i}]^H \in \mathbb{C}^{\ell_i}$, vector of received symbols during training phase

$$\mathbf{y}_i = h_i \mathbf{p} + \mathbf{w} \tag{1}$$

with $\mathbf{w} \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}_{\ell_i})$

Simplified Channel Model II

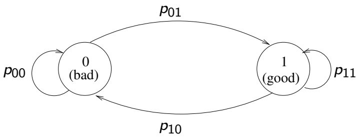

Figure: The Gilbert-Elliot channel model

- $h_i[k]$ can be written as

$$h_i[k] = x(h_0 - h_1) + \frac{1}{2}(h_0 + h_1), \tag{2}$$

- Let $\mathbf{v} \triangleq \frac{(h_0 - h_1)\mathbf{p}}{|h_0 - h_1| \|\mathbf{p}\|}$, and

$$\tilde{y} \triangleq \mathbf{v}^H \left[\mathbf{y} - \frac{1}{2}(h_0 + h_1)\mathbf{p} \right] = x\, |h_0 - h_1|\, \|\mathbf{p}\| + w, \tag{3}$$

where $w \sim \mathcal{CN}(0, \sigma^2)$.

Simplified Channel Model III

- Since $x \in \mathbb{R}$, $\Re\{\tilde{y}\}$ is sufficient to determine h.
- Applying the MAP decision rule

$$\Theta_i[k] = \begin{cases} 1, & \text{if } \lambda_i[k] \geq \eta_i \\ 0, & \text{otherwise}, \end{cases} \tag{4}$$

where

$$\lambda_i[k] \triangleq \ln \frac{P_{\ell_i}\left(\tilde{y}_i[k] | S_i[k] = 1\right)}{P_{\ell_i}\left(\tilde{y}_i[k] | S_i[k] = 0\right)} \tag{5}$$

$$= \frac{\sqrt{\ell_i E_p}\, |h_0 - h_1|\, \Re\{\tilde{y}\}}{\sigma^2/2}. \tag{6}$$

and

$$\eta_i \triangleq \ln \frac{P(s_i[k] = 0)}{P(s_i[k] = 1)} = \ln \frac{1 - p_{11}^{(i)}}{p_{01}^{(i)}}. \tag{7}$$

- If $\ell_i = 0$ is used for some i, then $\Theta_i = 1$ if $P(S_i = 1) \geq P(S_i = 0)$, and $\Theta_i = 0$ otherwise.

Sequence of events I

- At beginning of frame k, state of system transits to $\mathbf{S}[k] = [S_i[k]]_{i=1}^N$ according to $P(\mathbf{s}'|\mathbf{s})$
- Receiver decides on $\mathbf{l}[k] \in \mathcal{L}$ at beginning of frame k, where $\mathcal{L} \triangleq \left\{ \mathbf{l} : 1 \leq \ell_i \leq L, \sum_{i=1}^N \ell_i = N \right\}$
- Based on observation $\mathbf{\Theta}[k]$ from training phase, receiver selects antenna $n \in \mathcal{C}$where $\mathcal{C} \triangleq \{1, \ldots, N\}$
- Error check on data packet performed, resulting in observation $Z[k] \in \{0 \text{ (Error)}, 1 \text{ (No Error)}\}$

Sequence of events II

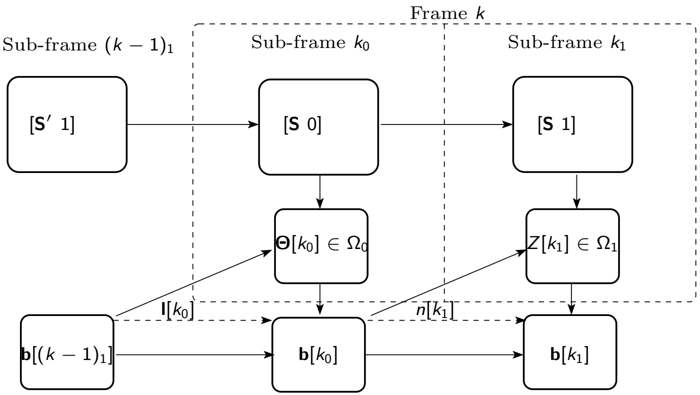

Figure: Sequence of events

Components of POMDP

- State Space $\mathcal{S} \triangleq \{0,1\}^{N+1}$, state $\mathbf{S}_m[k_m]$, $m=0$ denotes training period, $m=1$ denotes data packet reception period within a frame k
- Action Space $\mathcal{A} \triangleq \mathcal{L} \times \mathcal{C}$: Two parts:
  - Pilot allocation vector $\mathbf{l} = [\ell_i]_{i=1}^{N} \in \mathcal{L}$, where $\mathcal{L} \triangleq \left\{\mathbf{l} : \ell_i \in \{0,\ldots,L\} \forall i, \sum_{i=1}^{N} \ell_i = L\right\}$
  - Antenna selection decision $n \in \mathcal{C} \triangleq \{1,\ldots,N\}$
- Observation Space $\Omega \triangleq \Omega_0 \cup \Omega_1$: Also two parts:
  - Binary channel state observations at the antennas, $\mathbf{\Theta}[k_0] = [\Theta_i[k_0]]_{i=1}^{N} \in \Omega_0 \triangleq \{0,1\}^N$
  - Packet error indication $Z[k_1] \in \Omega_1 \triangleq \{0,1\}$

Components of POMDP (Contd.)

- Reward:
  - Given decision $\{\mathbf{l}[k_m], n[k_m]\}$, and $\mathbf{s}_m[k_m]$,

$$R[k_m] = m\mathbb{1}_{\{s_{m,n}=1\}} \tag{8}$$

  - Expected total discounted reward of POMDP over infinite horizon gives a measure of expected total number of bits that can be delivered
- Belief Vector: $\mathbf{b}[k_m]$
  - Component $b_{\mathbf{s}_m}[k_m] = P(\mathbf{s}_m | \text{dec. and obs. history}) \in [0, 1]$
- Policy:
  - π specifies the action to be taken at each decision point
  - Optimal policy at decision point k_m (end of decision period $k_m - 1$) maps the belief vector $\mathbf{b}[k_m - 1]$ to an action $A[k_m] = \{\mathbf{l}[k_m], n[k_m]\} \in \mathcal{A}$.

Objective

- Objective: Find π^*

$$\pi^* = \arg\max_{\pi} \mathbb{E}_{\pi} \left\{ \sum_{\{k_m = 1_0, 1_1, \ldots\}} \beta^q R[k_m] \Big| \mathbf{b}[0] \right\} \tag{9}$$

$\beta \in [0, 1)$, $q \triangleq 2(k-1) + m \ \forall k, m$

Value function I

- $V(\mathbf{b}[k_m])$, represents *maximum* expected discounted reward that can be obtained starting in the belief state $\mathbf{b}[k_m]$.
- Given action $A[k_m+1]$ and observation $o[k_m+1]$ reward accumulated starting from point k_m+1 consists of two parts:
  - the immediate reward $R[k_m+1] = m'z$, and
  - the maximum expected future reward $V(\mathbf{b}[k_m+1])$
- Optimality equations (Bellman Equations) can be written as:

$$\begin{aligned} V(\mathbf{b}[k_0]) &= \max_{A\in\mathcal{A}} \sum_{\mathbf{s}_0\in\mathcal{S}} b_{\mathbf{s}_0}[k_0] \sum_{z\in\Omega_1} P_A(z|\mathbf{b}[k_0]) \cdot \\ &\quad [z\cdot 1 + \beta V(f(\mathbf{b}[k_0], A, z))] \qquad (10) \\ V(\mathbf{b}[k_1]) &= \max_{A\in\mathcal{A}} \sum_{\mathbf{s}_1\in\mathcal{S}} b_{\mathbf{s}_1}[k_1] \cdot \\ &\quad \sum_{\theta\in\Omega_0} \beta P_A(\theta|\mathbf{b}[k_1]) V(f(\mathbf{b}[k_1], A, \theta)). \qquad (11) \end{aligned}$$

- Here, $\forall o \in \Omega_{m'}$, and $\forall A \in \mathcal{A}$,

$$P_A(o|\mathbf{b}[k_m]) = \sum_{\mathbf{s}'_{m'} \in \mathcal{S}} P_A\left(o|\mathbf{s}'_{m'}\right) \sum_{\mathbf{s}_m \in \mathcal{S}} b_{\mathbf{s}_m}[k_m] P(\mathbf{s}'_{m'}|\mathbf{s}_m) \qquad (12)$$

- For the simple channel model,

$$P_A\left(\Theta_i = 1 | S_{0,i} = s\right) = Q\left(\kappa_i \left(\frac{\eta_i}{\kappa_i^2} - x_i\right)\right) \qquad (13)$$

where $\kappa_i = |h_0 - h_1| \sqrt{\frac{2\ell_i E_p}{L\sigma^2}}$, and $x_i = -\frac{1}{2}$ if $s = 0$ and $x_i = +\frac{1}{2}$ if $s = 1$.

- Updated belief vector, $\mathbf{b}[k_m + 1]$ is obtained applying Bayes' rule, as

$$\begin{aligned} b_{\mathbf{s}'_{m'}}[k_m+1] &= P\left(\mathbf{S}_{m'}[k_m+1] = \mathbf{s}'_{m'} | \mathbf{b}[k_m], A, o\right) \\ &= \frac{\sum_{\mathbf{s}_m \in \mathcal{S}} b_{\mathbf{s}_m}[k_m] P(\mathbf{s}'_{m'}|\mathbf{s}_m) P_A(o|\mathbf{s}'_{m'})}{\sum_{\mathbf{s}'_{m'} \in \mathcal{S}} P_A(o|\mathbf{s}'_{m'}) \sum_{\mathbf{s}_m \in \mathcal{S}} b_{\mathbf{s}_m}[k_m] P(\mathbf{s}'_{m'}|\mathbf{s}_m)}. \end{aligned}$$

Value iteration

- Use 10 and 11 as assignment operation repeatedly, until value converges to V^*
- If the V^* can be computed, can be used directly in a greedy policy to get optimal behavior
- Greedy policy:

$$\begin{aligned}\pi(\mathbf{b}[k_m]) &= \arg\max_{A} \Bigg[\sum_{\mathbf{s}_m \in \mathcal{S}} b_{\mathbf{s}_m}[k_m] R[k_m] \\ &\quad + \beta \sum_{o \in \Omega_{m'}} P_A(o|\mathbf{b}[k_m]) V^*(\mathbf{b}[k_m+1]) \Bigg] \end{aligned} \tag{14}$$

- For finite horizon, V^* is piecewise linear and convex (PWLC)
- For infinite horizon, V^* is convex but not necessarily PWL
- $\therefore$ a PWL approximation is found and used

Algorithms

- Use PWL property of value function to represent it as finite set of vectors
- Exact Consider entire belief space
  Grow (Witness algorithm [Littman, 1994]), or
  Prune (Incremental Pruning [Cassandra et al., 1997]) set of vectors at each iteration
- Approximate Consider finite set of belief points
  (PBVI [Zhou and Hansen, 2001], SARSOP [Kurniawati et al., 2008], etc.)

Setup

- $N = 2$, $L = 4$
- Stationary probability of being in good state, $\bar{p}_1 = 0.5$
- Transition probability, $p_{01} = 0.2 \Rightarrow p_{11} = 0.8$
- POMDP solution compared to scheme with equal allocation $\ell_1 = \ell_2 = 2$ and greedy selection in every frame

Plot

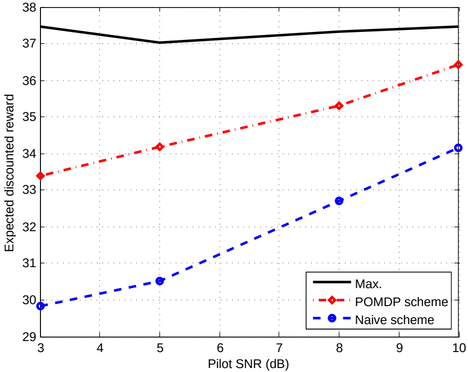

Figure: Performance plot with $N = 2,\ L = 4$

Conclusion and Future Work

- Problem of pilot allotment and selection modeled as a POMDP
- Performance of POMDP solution compared to that of a naive scheme
- Future work:
  - Consider effect of estimation error on packet error probability
  - Variations of problem

Thank

You